El Método de Stein y algunas Aplicaciones

Larry Goldstein

University of Southern California

Outline

I. Teoría basica

II. 2 Aplicaciones

El Lema de Stein

Caracterización de la distribución $\mathcal{N}(0, \sigma^2)$:

$$E[Zf(Z)] = \sigma^2 E f'(Z) \quad \text{si y sólo si } Z \sim \mathcal{N}(0, \sigma^2).$$

El Lema de Stein

Caracterización de la distribución $\mathcal{N}(0,\sigma^2)$:

$$E[Zf(Z)] = \sigma^2 Ef'(Z) \quad \text{si y sólo si } Z \sim \mathcal{N}(0,\sigma^2).$$

Tomamos la diferencia y la usamos en el lado izquierdo de una ecuación diferencial, junto a una función de prueba h en el lado derecho,

$$\sigma^2 f'(y) - yf(y) = h(y/\sigma) - Eh(Z).$$

El Lema de Stein

Caracterización de la distribución $\mathcal{N}(0, \sigma^2)$:

$$E[Zf(Z)] = \sigma^2 E f'(Z) \quad \text{si y sólo si } Z \sim \mathcal{N}(0, \sigma^2).$$

Tomamos la diferencia y la usamos en el lado izquierda de una ecuación diferencial, junto a una función de prueba h en el lado derecho,

$$\sigma^2 f'(y) - y f(y) = h(y/\sigma) - Eh(Z).$$

Para una Y dada, muestre que $Eh(Y/\sigma)$ es próxima a $Eh(Z)$, mostrando que la esperanza del lado izquierdo es pequeña, y que por lo tanto Y es aproximadamente normal.

El Método de Stein

Una colección de maneras de demostrar que

$$E[\sigma^2 f'(Y) - Yf(Y)]$$

es pequeña. Una lista, muy incompleta:

1. Pares Intercambiables
2. Apareamientos de sesgo de tamaño, [si $X \sim dF(x)$ tenemos $X^s \sim xdF(x)/EX$]
3. Apareamientos de sesgo cero.

La Distribución Y-Sesgo Cero

Para toda Y con media cero y varianza σ^2, existe una, y sólo una, distribución $\mathcal{L}(Y^*)$ tal que

$$E[Yf(Y)] = \sigma^2 Ef'(Y^*)$$

para toda f suave (Goldstein y Reinert, 1997). La distribución de Y^* es conocida como la distribución Y-sesgo cero.

El Lema de Stein se convierte en: La transformación de distribuciones $Y \rightarrow Y^*$ tiene a la $\mathcal{N}(0, \sigma^2)$ como su único punto fijo.

Cota en L^1

Principio: Si $\mathcal{L}(Y)$ y $\mathcal{L}(Y^*)$ están cerca, $\mathcal{L}(Y)$ está cerca del punto fijo de la transformación, y, por lo tanto, cerca del único punto fijo, la Normal.

En la distancia L^1 (Wasserstein, Dudley, o Kantarovich), $\mathrm{Var}(Y) = 1$, este principio se manifiesta en

$$||\mathcal{L}(Y) - \mathcal{L}(Z)|| \leq 2||\mathcal{L}(Y^*) - \mathcal{L}(Y)||.$$

A veces podemos calcular el lado derecho, y mostrar que es pequeño, usando apareamiento.

Apareamientos de sesgo cero

Sustituyendo Y y tomando la esperanza en la ecuación de Stein

$$h(y/\sigma) - Eh(Z) = \sigma^2 f'(y) - yf(y)$$

da

$$\begin{aligned} Eh(Y/\sigma) - Eh(Z) &= E[\sigma^2 f'(Y) - Yf(Y)] \\ &= \sigma^2 E[f'(Y) - f'(Y^*)]. \end{aligned}$$

Entonces, que Y esté cerca de Y^* implica que Y esta cerca de la Normal. Ahora podemos usar cotas en la ecuación de Stein, como $||f''||_\infty \leq 2||h'||_\infty$, para hacer este agrumento preciso.

Ejemplos

Outline

I. Teoría basica

II. 2 Aplicaciones

   a. Sumas Independientes

   b. Estructuras Jerárquicas

Sesgo de Tamaño y Sesgo Cero

Sea Y no negativa con media finita, $EY = \mu$. Recordemos que Y^s tiene la distribución Y-sesgo de tamaño si

$$dF^s(y) = ydF(y)/\mu \quad \text{o} \quad E[Yf(Y)] = \mu Ef(Y^s)$$

para toda f suave. Sesgo de tamaño es responsable de la paradoja de tiempos de esperanza, y de ciertos tipos de errores en muestras.

Para obtener la transformación de sesgo cero (a partir de la transformación de sesgo de tamaño) reemplazamos la media por la varianza, y f por f'.

Dando Sesgo Cero (de Tamaño) a una Suma Independiente

Para dar sesgo cero (de tamaño) a una suma

$$Y = \sum_{i=1}^{n} X_i$$

de variables independientes con media cero (no negativas), escoja una en proporción con su propia varianza (media) y reemplácela con su versión con sesgo.

Esto da la intuición a una pregunta básica: Cúando es una suma Y próxima a la Normal?

Racional del TCL para el Sesgo Cero

Si Y es una suma de variables comparables, independientes, y con media cero, Y^* difiere de Y en sólo un 'típico' sumando.

Entonces, cuando hay muchas variables, Y^* está cerca de Y, por lo que Y es casi el punto fijo de la tranformación de sesgo cero, y por lo tanto próxima de la Normal.

Sumas de Variables Independientes e Idénticamente Distribuidas

Cuando $Y = n^{-1/2} \sum X_i$ con $X, X_1, \ldots, X_n$ i.i.d. G con media cero, varianza 1, $P(I = i) = 1/n$,

$$Y^* - Y = n^{-1/2}(X_I^* - X_I).$$

Entonces, por la cota anterior

$$||F - \Phi|| \leq \frac{2}{\sqrt{n}} ||G^* - G||.$$

Se puede mostrar que la función de distribución G^* de X^* es dada explícitamente por

$$G^*(x) = E[X(X - x)\mathbf{1}(X \leq x)].$$

Constantes de Berry Esseen, Dependiendo de la Distribución

Bernoulli* es Uniforme, y

$$||F - \Phi|| \leq c\frac{E|X_1|^3}{\sqrt{n}}, \quad \text{donde } c = 1.$$

Para $\mathcal{U}[-\sqrt{3}, \sqrt{3}]$, tenemos

$$G^*(x) = -\frac{\sqrt{3}x^3}{36} + \frac{\sqrt{3}x}{4} + \frac{1}{2} \quad \text{cuando} \quad x \in [-\sqrt{3}, \sqrt{3}],$$

nos da

$$||F - \Phi|| \leq \frac{\sqrt{3}}{4\sqrt{n}} = c\frac{E|X_1|^3}{3\sqrt{n}}, \quad \text{donde } c = 1/3$$

Ejemplos

Outline

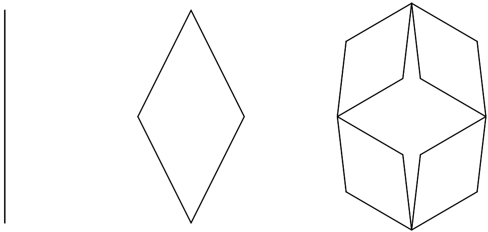
I. Teoría basica

II. 2 Aplicaciones

   a. Sumas Independientes

   b. Estructuras Jerárquicas

El Reticulado Diamante: Tres Etapas



Estructuras Jerárquicas

Propiedades (como conductancia) de un material a nivel 2 dependen de las propiedades a nivel 1, ...

$$X_2 = F(\mathbf{X}_1), \qquad X_1 = F(\mathbf{X}_0) \ldots$$

Modelos Jerárquicos

$$X_{n+1} = F(\mathbf{X}_n) \quad \text{donde} \quad \mathbf{X}_n = (X_{n,1}, \ldots, X_{n,k})^\mathsf{T}$$

con $X_{n,i}$ independientes, cada una con distribución X_n.

Condiciones en F, dadas por Li y Rogers, y Wehr, implican la ley débil (que aquí asumimos)

$$X_n \to_p c,$$

y las dadas por Woo y Wehr implican

$$W_n \to_d \mathcal{N}(0,1), \quad \text{para} \quad W_n = \frac{X_n - EX_n}{\sqrt{\mathsf{Var}(X_n)}}.$$

El TCL Clásico, Interpretado como un Modelo Jerárquico

Tomando F para darnos la *media*

$$F(x_1, x_2) = \frac{x_1 + x_2}{2}$$

da en distribución

$$X_n = \frac{X_{0,1} + \cdots + X_{0,2^n}}{2^n}.$$

En la etapa n hay $N = 2^n$ variables, por lo que esperaríamos una cota a la Normal Z de la forma

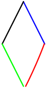
$$||W_n - Z|| \le C\gamma^n \quad \text{donde} \quad \gamma^n = N^{-1/2} = (1/\sqrt{2})^n.$$

Funciones Promedio

Decimos que F es una función (estríctamente) promedio cuando

1. $\min_i x_i \leq F(\mathbf{x}) \leq \max_i x_i$, y estríctamente cuando $\min_i x_i < \max_i x_i$.
2. $F(\mathbf{x}) \leq F(\mathbf{y})$ cada vez que $x_i \leq y_i$, y estríctamente cuando $x_j < y_j$ para alguna j.

Decimos F es una función (estríctamente) promedio normalizada, si $F(\mathbf{x})/F(\mathbf{1}_k)$ es (estríctamente) promedio, donde $\mathbf{1}_k = (1, \ldots, 1)$.



La Función da Conductividad para el Reticulado Diamante

Reglas para combinar conductividades ($= 1/$(resistencias)) en paralelo y en serie

$$L_1(x_1, x_2) = x_1 + x_2, \quad L_{-1}(x_1, x_2) = (x_1^{-1} + x_2^{-1})^{-1}$$

dan la función da conductividad ponderada (con pesos $w_i > 0$) para el reticulado diamante

$$F(\mathbf{x}) = \left(\frac{1}{w_1 x_1} + \frac{1}{w_2 x_2} \right)^{-1} + \left(\frac{1}{w_3 x_3} + \frac{1}{w_4 x_4} \right)^{-1},$$

una función estríctamente promedio normalizada.

Recursión Lineal Aproximada

Escribimos $X_{n+1} = F(\mathbf{X}_n)$ como una recursión lineal con (‘pequeña’) perturbación R_n,

$$X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n, \quad n \geq 0,$$

donde $c_n = EX_n$, $\boldsymbol{\alpha}_n = F'(\mathbf{c}_n)$, $\mathbf{c}_n = (c_n, \ldots, c_n)^\mathsf{T} \in \mathbb{R}^k$, y F' es el gradiente de F.

Asumiendo que el gradiente $\boldsymbol{\alpha} = F'(\mathbf{c})$ en el limite c no es una múltiplo escalar de un vector de la base canónica (por ejemplo $(1, 0, \ldots)$), evita casos triviales como $F(x_1, x_2) = x_1$ e implica que $\lambda = ||\boldsymbol{\alpha}|| < 1$ cuando F es promedio.

Función de Contracción en L^1

Dado $\boldsymbol{\alpha} \in \mathbb{R}^k$ con $\lambda = ||\boldsymbol{\alpha}|| \neq 0$, sea

$$Y = \sum_{i=1}^{k} \frac{\alpha_i}{\lambda} W_i,$$

donde W_i son variables independientes con media cero, varianza uno, distribuidas como W. Entonces

$$||Y - Y^*|| \leq \varphi \, ||W - W^*||,$$

y $\varphi = \sum_i |\alpha_i|^3 / (\sum_i \alpha_i^2)^{3/2} < 1$ si y sólo si $\boldsymbol{\alpha}$ no es un múltiplo escalar de un vector de la base canónica.

Iteración Lineal

Normalizando $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n$ para dar W_{n+1}, con $\lambda_n = ||\boldsymbol{\alpha}_n||, EX_n = c_n$ y $\sigma_n^2 = \mathsf{Var}(X_n)$, tenemos

$$W_{n+1} = \sum_{i=1}^{k} \frac{\alpha_{n,i}}{\lambda_n} W_{n,i} \quad \text{con} \quad W_n = \frac{X_n - c_n}{\sigma_n}.$$

Iterando la contracción da

$$||W_n - Z|| \leq 2||W_n - W_n^*|| \leq 2 \left(\prod_{i=0}^{n-1} \varphi_i \right) ||W_0 - W_0^*||.$$

Iteración no Lineal

Sea $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n$, donde $\mathbf{X}_n$ es un vector de variables i.i.d. con la misma distribución que X_n, $EX_n = c_n$, $\mathrm{Var}(X_n) = \sigma_n^2$, y $\lambda_n = ||\boldsymbol{\alpha}_n|| \neq 0$. Fije

$$Y_n = \sum_{i=1}^{k} \frac{\alpha_{n,i}}{\lambda_n} W_{n,i} \quad \text{donde} \quad W_n = \frac{X_n - c_n}{\sigma_n}$$

y, midiendo la discrepancia respecto de linealidad,

$$\beta_n = E|W_{n+1} - Y_n| + \frac{1}{2} E|W_{n+1}^3 - Y_n^3|;$$

además, sea

$$\varphi_n = \sum_i |\alpha_{n,i}|^3 / (\sum_i \alpha_{n,i}^2)^{3/2}.$$

Teorema 1: Para $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n$, si existe $(\beta, \varphi) \in (0,1)^2$ tal que

$$\limsup_{n\to\infty} \frac{\beta_n}{\beta^n} < \infty \quad \text{y} \quad \limsup_{n\to\infty} \varphi_n = \varphi,$$

entonces, tomando $\gamma = \beta$ cuando $\varphi < \beta$, o tomando $\gamma \in (\varphi, 1)$ cuando $\beta \le \varphi$, existe C tal que

$$||W_n - Z|| \le C\gamma^n.$$

Teorema 1: Para $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n$, si existe $(\beta, \varphi) \in (0,1)^2$ tal que

$$\limsup_{n\to\infty} \frac{\beta_n}{\beta^n} < \infty \quad \text{y} \quad \limsup_{n\to\infty} \varphi_n = \varphi,$$

entonces, tomando $\gamma = \beta$ cuando $\varphi < \beta$, o tomando $\gamma \in (\varphi, 1)$ cuando $\beta \leq \varphi$, existe C tal que

$$||W_n - Z|| \leq C\gamma^n.$$

Ahora aplique Teorema 1 a secuencias generadas por funciones promedio F.

Teorema 2 Sea X_0 una variable aleatoria, non constante, con $P(X_0 \in [a,b]) = 1$ y $X_{n+1} = F(\mathbf{X}_n)$ donde $F : [a,b]^k \to [a,b]$, en dos veces contínuamente diferenciable. Suponga que F es promedio y que $X_n \to_p c$, con $\boldsymbol{\alpha} = F'(\mathbf{c})$ no un múltiplo escalar de un vector de la base canónica. Entonces, con Z una variable $\mathcal{N}(0,1)$, para todo $\gamma \in (\varphi, 1)$ existe C tal que

$$||W_n - Z|| \leq C\gamma^n \quad \text{donde} \quad \varphi = \frac{\sum_{i=1}^k |\alpha_i|^3}{(\sum_{i=1}^k |\alpha_i|^2)^{3/2}},$$

es un número positivo, estríctamente menor que 1. El valor φ alcanza su mínimo de $1/\sqrt{k}$ si y sólo si los componentes de $\boldsymbol{\alpha}$ son iguales.

Convergencia Rápida para el Reticulado Diamante

Defina la 'red con pesos iguales en cada lado' como la red con $\mathbf{w} = (w, w, 2-w, 2-w)^{\mathsf{T}}$ con $w \in (1,2)$; tales pesos son positivos, y satisfacen $F(\mathbf{w}) = 1$.

Cuando $w = 1$ todos los pesos son iguales, y tenemos $\boldsymbol{\alpha} = 4^{-1}\mathbf{1}_4$, por lo tanto φ logra su valor mínimo $1/2 = 1/\sqrt{k}$, a la velocidad (de convergencia) $N^{-1/2+\epsilon}$ ($\gamma^n = N^{-\theta} = 4^{-n\theta}$).

Para $1 \leq w < 2$ tenemos $1/2 \leq \varphi < 1/\sqrt{2}$, y cuando $w \uparrow 2$ la red con pesos iguales en cada lado alcanza su rango de convergencia menos favorable, $N^{-1/4+\epsilon}$.

Convergencia Lenta para el Reticulado Diamante

Con sólo la restricción de que los pesos sean positivos y satisfagan $F(\mathbf{w}) = 1$ considere, para $t > 0$,

$$\mathbf{w} = (1 + 1/t, s, t, 1/t)^{\mathsf{T}} \quad \text{donde}$$

$$s = [(1 - (1/t + t)^{-1})^{-1} - (1 + 1/t)^{-1}]^{-1}.$$

Cuando $t = 1$ tenemos $s = 1$ y $\varphi = 11\sqrt{2}/27$.

Cuando $t \to \infty$, $\boldsymbol{\alpha}$ tiende al vector $(1, 0, 0, 0)$, por lo que $\varphi \to 1$.

Como $11\sqrt{2}/27 < 1/\sqrt{2}$, γ asume todos los valores en $(1/2, 1)$, correspondiendo a $N^{-\theta}$ para todo $\theta \in (0, 1/2)$.

Resumen

- Sumas Independientes
  - Reemplace Uno
- Estructuras Jerárquicas
  - Contracción Iterada Aproximada

Más Ejemplos

Outline

I. Teoría basica

II. 4 Aplicaciones

   a. Sumas Independientes
   b. Estructuras Jerárquicas
   c. Teorema Central del Límite Combinatorio
   d. Medida Cónica sobre la Esfera

Teorema Central del Límite Combinatorio

Dada una matriz real A, de $n \times n$, obtenga cotas a la aproximación Normal para

$$Y = \sum_{i=1}^{n} a_{i,\pi(i)},$$

donde π es una permutación aleatoria en $\mathcal{S}_n$, el grupo simétrico.

(Hoeffding, Chen y Ho, Bolthausen, von Bahr)

Distribución Uniforme $\mathcal{U}(\mathcal{S}_n)$

1. Muestreo aleatorio simple es un caso especial del caso especial donde $a_{ij} = c_i d_j$.
2. Medida de uniformidad de π, tomando $a_{ij} = |i - j|$ da $a_{i,\pi(i)} = |i - \pi(i)|$. $Y = 0$ cuando $\pi =$ id; ¿Qué tan lejos está Y de cero cuando π es uniforme?
3. Distribución de la estadística del test de permutación, por ejemplo, para determinar la concordancia entre X_i e Y_i, tome $a_{ij} = |X_i - Y_j|$ para determinar que tal bien $\sum_i |X_i - Y_i|$ se ajusta a la distribución de

$$\sum_i |X_i - Y_{\pi(i)}|.$$

Pares Intercambiables de Stein

Un par de variables intercambiables (Y, Y') que satisface

$$E(Y'|Y) = (1-\lambda)Y, \quad \lambda \in (0,1).$$

Podemos aplicar el método para un Y dado si podemos construir (Y, Y') que sea un par intercambiable de Stein.

Sesgo Cero a partir de un Par Intercambiable

Sea (Y', Y'') un par de Stein con distribución conjunta $F(y', y'')$ y $\mathrm{Var}(Y') = \sigma^2$. Sea

$$dF^\dagger(y^\dagger, y^\ddagger) = \frac{(y^\dagger - y^\ddagger)^2}{2\lambda\sigma^2} dF(y^\dagger, y^\ddagger),$$

y $U \sim \mathcal{U}[0,1]$ independiente. Entonces, cuando $(Y^\dagger, Y^\ddagger) \sim F^\dagger$,

$$Y^* = UY^\dagger + (1-U)Y^\ddagger$$

tiene distribución Y-sesgo cero.

TCL Combinatorio, $Y' = \sum_{i=1}^{n} a_{i,\pi'(i)}$

Dada $\pi' \sim \mathcal{U}(\mathcal{S}_n)$, sea τ_{IJ} la transposición de I y J, escogidas distintas y uniformemente, y sea

$$\pi'' = \pi' \tau_{I,J}.$$

Entonces, con Y'' construida usando π'', (Y', Y'') es un par de Stein que satisface

$$E(Y''|Y') = (1 - \frac{2}{n-1})Y'.$$

Además,

$$Y'' - Y' = a_{I,\pi'(J)} + a_{J,\pi'(I)} - (a_{I,\pi'(I)} + a_{J,\pi'(J)}).$$

Sesgo del Cuadrado de la Diferencia y Apareamientos

Con π dada, para construir $\pi^\dagger, \pi^\ddagger$ en el mismo espacio, considere $I^\dagger, J^\dagger, K^\dagger, L^\dagger$ con distribución

$$p(i,j,k,l) = \frac{\left[(a_{ik} + a_{jl}) - (a_{il} + a_{jk})\right]^2}{4n^2(n-1)\sigma^2}.$$

Ahora (con $\pi = \pi'$) defina

$$\pi^\dagger = \begin{cases} \pi\tau_{\pi^{-1}(K^\dagger),J^\dagger} & \text{si } L^\dagger = \pi(I^\dagger), K^\dagger \neq \pi(J^\dagger) \\ \pi\tau_{\pi^{-1}(L^\dagger),I^\dagger} & \text{si } L^\dagger \neq \pi(I^\dagger), K^\dagger = \pi(J^\dagger) \\ \pi\tau_{\pi^{-1}(K^\dagger),I^\dagger}\tau_{\pi^{-1}(L^\dagger),J^\dagger} & \text{de lo contrario,} \end{cases}$$

y $\pi^\ddagger = \pi^\dagger\tau_{I^\dagger,J^\dagger}$. Tenemos $Y^* = UY^\dagger + (1-U)Y^\ddagger$.

Cota a travez del cálculo de $E|Y^* - Y'|$

Teorema 3 Definiendo

$$a_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij} \quad a_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n} a_{ij} \quad a_{\bullet\bullet} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij}$$

$$a_3 = \sum_{i,j=1}^{n} |a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}|^3,$$

con $\sigma^2 = \mathrm{Var}(Y')$, obtenemos

$$||F - \Phi|| \leq \frac{a_3}{(n-1)\sigma^3}\left(16 + \frac{56}{n-1} + \frac{8}{(n-1)^2}\right).$$

Ejemplos

Outline

I. Teoría basica

II. 4 Aplicaciones

   a. Sumas Independientes
   b. Estructuras Jerárquicas
   c. Teorema Central del Límite Combinatorio
   d. Medida Cónica sobre la Esfera

Medida Cónica $\mathcal{C}_p^n$ en $\mathbb{R}^n$

Estudiamos proyecciones de medidas sobre cuerpos convexos en $\mathbb{R}^n$ para un ejemplo específico.

Medida Cónica $\mathcal{C}_p^n$ en $\mathbb{R}^n$

$$S(\ell_p^n) = \{\mathbf{x} : \sum_{i=1}^{n} |x_i|^p = 1\}, \quad B(\ell_p^n) = \{\mathbf{x} : \sum_{i=1}^{n} |x_i|^p \leq 1\}$$

Con μ^n la medida Lebesgue en $\mathbf{R}^n$, para $A \subset S(\ell_p^n)$ y

$$[0,1]A = \{ta : a \in A, 0 \leq t \leq 1\}$$

sea

$$\mathcal{C}_p^n(A) = \frac{\mu^n([0,1]A)}{\mu^n(B(\ell_p^n))}.$$

Medida Cónica $\mathcal{C}_p^n$ en $\mathbb{R}^n$

Casos especiales

1. $p = 1$: La Distribución Uniforme sobre el simplex

$$\sum_{i=1}^{n} |x_i| = 1.$$

2. $p = 2$: La Distribución Uniforme sobre la esfera

$$\sum_{i=1}^{n} x_i^2 = 1.$$

Proyección

Para $\mathbf{X} \sim \mathcal{C}_p^n$, para algún $p > 0$ y $\boldsymbol{\theta} \in \mathbb{R}^n$ un vector unitario, considere la proyección

$$Y = \boldsymbol{\theta} \cdot \mathbf{X}.$$

Cuando $\boldsymbol{\theta} = n^{-1/2}(1, 1, \ldots, 1)$, tenemos $Y = n^{-1/2} \sum_{i=1}^{n} X_i$.

Diaconis y Freedman: para $p = 2$ consideraron cotas de variación total.

Meckes y Meckes: para vectores aleatorios con simetrías generales, consideraron cotas en las distancias supremo y de variación total.

Construcción de Sesgo Cero para Vectores con Componentes Simétricas

$$(Y_1, \ldots, Y_n) =_d (e_1 Y_1, \ldots, e_n Y_n), \quad \forall e_i \in \{-1, 1\}.$$

Como $Y_i =_d -Y_i$ y $(Y_i, Y_j) =_d (Y_i, -Y_j)$ cuando $i \neq j$, cuando existen momentos de segundo order tenemos

$$EY_i = 0 \quad \text{y} \quad \mathrm{Cov}(Y_i, Y_j) = 0.$$

Con $\sigma_i^2 = \mathrm{Var}(Y_i)$, la construcción depende de las distribuciones *sesgo cuadrado* en la dirección i,

$$EY_i^2 f(\mathbf{Y}) = \sigma_i^2 E f(\mathbf{Y}^i) \quad \text{o} \quad dF^i(\mathbf{y}) = \frac{y_i^2}{\sigma_i^2} dF(\mathbf{y}).$$

De Sesgo Cuadrado a Sesgo Cero

Sea $Y = \sum_{i=1}^{n} Y_i$, I un índice aleatorio independiente con distribución $P(I = i) \propto \sigma_i^2$ y $U \sim \mathcal{U}[-1, 1]$ independiente de las otras variables. Entonces

$$Y^* = UY_I^I + \sum_{j \neq I} Y_j^I.$$

Bajo simetría, generaliza la construcción de 'reemplace una' para variables independientes, dada antes.

Para apareamientos bajo dependencia, escoja i de acuerdo con I, genere y_i^i, luego 'ajuste' $Y_j, j \neq i$ de acuerdo a la distribución condicional dada $Y_i = y_i^i$.

Representación de la Medida Cónica

Si $\{G_j, \epsilon_j, j = 1, \ldots, n\}$ son variables independientes con $G_j \sim \Gamma(1/p, 1)$, $G_{1,n} = \sum_{i=1}^n G_i$ y $\epsilon_j \in \{-1, 1\}$ equiprobables, tenemos

$$\mathbf{X} = \left(\epsilon_1 (\frac{G_1}{G_{1,n}})^{1/p}, \ldots, \epsilon_n (\frac{G_n}{G_{1,n}})^{1/p} \right) \sim \mathcal{C}_p^n.$$

En particular, esta representación confirma que la medida cónica es simétrica con respecto a sus componentes.

Sesgo Cuadrado en una Dirección Dada

Con $G_i' \sim \Gamma(2/p, 1)$ independiente,

$$X_i^i = \epsilon_i \left(\frac{G_i + G_i'}{G_{1,n} + G_i'} \right)^{1/p}$$

tiene la distribución de sesgo cuadrado de X_i, y el vector (renormalizado) $\mathbf{X}^i$ con componentes

$$X_j^i = \begin{cases} \left(\frac{1-|X_i^i|^p}{1-|X_i|^p}\right)^{1/p} X_j & j \neq i \\ X_i^i & j = i \end{cases}$$

tiene la distribución $\mathbf{X}$ con sesgo cuadrado en la dirección i.

De $\mathbf{X}^i$ a Y^*

Ahora tenemos $\mathbf{X}^i$ para todo $i = 1, \ldots, n$. Recuerde que

$$Y = \boldsymbol{\theta} \cdot \mathbf{X} = \sum_{i=1}^{n} \theta_i X_i = \sum_{i=1}^{n} Y_i.$$

Hecho: El vector

$$\mathbf{Y}^i = (\theta_1 X_1^i, \ldots, \theta_n X_n^i)$$

tiene la distribución de $\mathbf{Y}$, con sesgo cuadro en la dirección i. Ahora defina $P(I = i) \propto \sigma_i^2 = \theta_i^2$ y

$$Y^* = U Y_I^I + \sum_{j \neq I} Y_j^I.$$

Cota L^1 para la aproximación Normal de (proyecciones de) la Medida Cónica

Theorem 4 Sea $\mathbf{X}$ una variable aleatoria con la medida cónica $\mathcal{C}_p^n$ en la esfera $S(\ell_p^n)$ para algún $p > 0$, y sea

$$Y = \sum_{i=1}^{n} \theta_i X_i$$

la proyeccíon en una dimensión de $\mathbf{X}$ a lo largo de la dirección $\boldsymbol{\theta} \in \mathbf{R}^n$ con $||\boldsymbol{\theta}|| = 1$. Entonces, con $\sigma_{n,p}^2 = \mathsf{Var}(X_1)$ y $m_{n,p} = E|X_1^1|$,

$$||F - \Phi|| \leq 3 \left(\frac{m_{n,p}}{\sigma_{n,p}} \right) \sum_{i=1}^{n} |\theta_i|^3 + \left(\frac{1}{p} \vee 1 \right) \frac{4}{n+2}.$$

Casos Especiales

$p = 1$

$$||F - \Phi|| \leq \frac{9}{\sqrt{2}} \sum_{i=1}^{n} |\theta_i|^3 + \frac{4}{n+2}$$

$p = 2$

$$||F - \Phi|| \leq \frac{9}{\sqrt{3}} \sum_{i=1}^{n} |\theta_i|^3 + \frac{4}{n+2}$$

Resumen

- Sumas Independientes
  - Reemplace Uno
- Estructuras Jerárquicas
  - Contracción Iterada Aproximada
- Teorema Central del Límite Combinatorio
  - Pares Intercambiables
- Medida Cónica Sobre la Esfera
  - Sesgo Cuadrado bajo Simetría